CptS 515 Advanced Algorithms FINAL ASSIGNMENT :

Sharath Kumar Karnati

DECEMBER 2023

QUESTION

Many people think that the dynamics of the stock market is not even predictable. However, after closely looking at a chart, you would agree that the trend of a stock should be predictable; at least, the price movement isnt as random as we thought (see an example chart on page 3). Nowdays, extremely short term price predication is a mature technology, pioneered by mathematician Jim Simons who successfully developed an algorithm in 1980s in predicting stock prices. Nowadays, 70 percent of stock trades in the United States are done by algorithms and hence, as a computer science graduate student, you shall understand how algorithms are, profitably, applied in the real world. Surprisingly, a stock chart has a very simple data structure: an array [1...k], for some large k. Each element [i] is the market data for the stock of day i, which is represented by a tuple of five numbers:

• open price of the day;
• close price of the day;
• highest price of the day;
• lowest price of the day; and
• number of shares traded in the day (called volume).

Notice that when i = k, day i refers to today. For major stocks traded in the United States, the charts are publically available for download; most of them are with a history of more than 2000 trading days (10 years); i.e., with the k 2000. In this exam, you are going to write a mini paper to address the following two problems:

• A. (30 percent) Given a stock chart a with k $\geq$ 2000, design one or more algorithms to predict the price for the next day. I will grade on efficiency, depth and correctness of your ideas.

• B. (30percent) Given two stock charts and , design one or more algorithms to measure the similarity between the two. I will grade on efficiency, depth and correctness of your ideas. The remaining 40 percent contributes to the quality and clarity of your writing. I shall emphasize again that your algorithms shall only take input as the stock chart(s) given.


12/4/22, 3:52 PM
Print SharpCharts from StockCharts.com
AMD Advanced Micro Devices, Inc. Nasdaq GS
2-Dec-2022
©StockCharts.com
Open 75.05 High 76.28 Low 73.65 Close 74.98 Volume 64.4M Chg -2.50 (-3.23%)
AMD (Daily) 74.98
Volume 64,376,596
110.0
107.5
105.0
102.5
100.0
97.5
95.0
92.5
90.0
87.5
85.0
82.5
80.0
77.5
74.98
72.5
70.0
67.5
65.0
62.5
60.0
55.0
250M
200M
150M
100M
50M
18 25 May 9 16 23 Jun6 13 21 27 Jul 11 18 25 Aug 8 15 22 29Sep 12 19 26 Oct 10 17 24 Nov 7 14 21 28Dec

Prediction of Stock Market prices and Finding the similarities between two stock charts

SHARATH KUMAR KARNATI, 011852253
ADVANCED ALGORITHMS
SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
WASHINGTON STATE UNIVERSITY
PULLMAN, UNITED STATES OF AMERICA
Email: s.karnati@wsu.edu

***Abstract*—Nowadays, a growing number of individuals are keenly interested in monitoring stock market trends. Thanks to technological advancements, we now have the capability to forecast stock market trends based on the historical market behaviour. Numerous companies are actively engaged in developeing tools and methods for this purpose. Predicting stock market trends is often facilitated by the use of algorithms, which streamline the process by providing a standardized approach applicable to any stock chart, rather than requiring a unique approach for each day's market data. This paper explores various algorithms employed in stock market forecasting and the exploration of similarities between two different stock charts.**

***Index Terms*—Stock Charts, Continuous Hidden Markov model,De-Bruijn's Graph, Maximum Posterior Estimate, Time series, Correlation, Dynamic Time Wrapping, Minimum Spanning Tree, Rectangular Granulation Information, Markov Chains**

I. Introduction

The stock market is a dynamic and vital financial marketplace that allows individuals to buy, sell, and issue shares of publicly traded companies. Not only does it serve as a platform for businesses to raise funds by issuing stocks, but it also provides a space for investors to purchase and trade these stocks. The ultimate goal of investing in the stock market can vary, whether it be for dividends, capital gains, or a combination of both. However, regardless of the personal objectives, the stock market plays a crucial role in the economy by facilitating capital allocation and providing businesses with a means of growth and expansion.

The stock market is not linear everyday. It always changes and is not consistent and also has an ability to change rapidly due to some unforeseen situations. So, it always becomes a chalenging task to predict the future behaviour by taking the historical data due to its very non-linear and fluctuating nature. To design a algorithm we have to be very careful and precise with the inputs we take and the algorithms that we choose.

Because the stock market is so volatile and dynamic, predicting stock values is extremely difficult. Data from stock charts can be analysed to find patterns that are not coincedental, indicating the possibility of predictability with close monitoring. When a firm has a lot of data, it can be used to predict its future and use algorithms to help with stock purchase or sale decisions. Effective strategies to forecast stock prices and improve investor returns are being intensively investigated by researchers.

A number of variables, including interest rates, ownerships shifts, mergers, and equity swings, can affect how volatile stock prices are. Fuzzy logic and Support Vector Machines(SVM's) are two examples of artificial intelligence (AI) approaches that have been integrated for stock price prediction in recent efforts. Even while these strategies have some potential characteristics, it's essential to acknowledge their inherit limitations.

In this paper, I am gonna propose some methods that can be used for the two given queries.

- 1. Using historical Stock chart's data, prediction of its data for the next upcoming day.
- 2. Similarities between two stock chart's α and β.

Now we are gonna introduce different types of models that can be used to solve the above two queries. We have given a Stock Chart which contains five data elements that are Openprice, Closeprice, Volume, High and Low which is represented with a tuple of 5 elements. We are gonna use these elements exclusively as the input for our models. We are gonna see various methods which includes time series, Markov Chains, Hidden Markov Chains, Dynamic Time Wraping,...etc. We have a data which is in frequency type , we have to first convert this to time series in every model as we are not using any machine Learning techniques to process the graph data. After conversion we are gonna employ various methodologies to model the algorithm that can be used to solve the above two queries.

II. FORECASTING THE STOCK PRICES :

A. USING THE CONTINUOUS-TIME HIDDEN MARKOV MODELS (CHMM's) and MAXIMUM A POSTERIOR ESTIMATE(MAP) :

We are going to use Continuous-time Hidden Markov's Model to determine the forcasted stock price. A continuous-time Hidden Markov's Model is simply an Hidden Markov Model(HMM) where we can see both the transitions between the hidden states and the observations arrival will be occured at a continuous time. A HMM is a generally a statistical

model which will be helpful to describe the evolution of observable events which are depended on some internal factors that cannot be observed directly but are hidden and effects the events.HMM's are generally used because of their good statistical basis, they are good with handling the fresh data, they are easier to develop and to analize effectively than compared to the existing models and also they will produce some identical patterns.

1) The APPROACH :

- Firstly, we are gonna use the continuous hidden Markov's model to do the modelling of the stock chart's data into a time series.
- Here a HMM will be denoted by a λ .

$$\lambda = (\pi, A, B)$$

Here,
A = Transition Matrix , this consists of the elements that gives us the probability of the transistion of one state into another state.
B = Emission Matrix, this gives us the $b_j(O_t)$ which represents the probability of observation of the sequence O = O_1,O_2,.....O_t when it is in state j.
pi represents the initial probabilities for the states when they are at t=1

- In a CHMM's, B is stated as

$$b_j(O_t) = \sum_{m=1}^{M} C_{jm} N(O_t, \mu_{jm}, \sigma_{jm})$$

Here,
M = Number of Gaussian mixture components
C_{jm} = The weight of m^{th} component mixture in the state j
μ_{jm} = Mean vector of the m^{th} component in the state j
$N(O_t, \mu_{jm}, \sigma_{jm})$ = The probability of the observation of the O_t in a Multidimensional Gaussian distribution.

- Now, as specified in the question let's take the O_t as of tuple which contains 5 elements that are [Openprice, Closeprice, High, Low, Volume]
- Now, we will represent these tuples as vector to define the observations of our HMM.

$$O_t = (\frac{Close - Open}{Open}, \frac{High - Open}{Open}, \frac{Open - Low}{Open})$$

- Therefore, now O_t = (DiffFrac,HighFrac,LowFrac)
As we are using values in fractions instead of using them in direct form we can achieve higher accuracy levels in forecasting.
- Now, we will compute the probability for the events and it is caluclated as

$$P(O|\lambda) = \Sigma P(O|Q, \lambda) P(Q, \lambda)$$

,
Here, Q = sequence of states $(q_1, q_2, ...q_t)$

- From the given, the number of days is k which will be k$\geq$2000.
- For this model, we are gonna equip a Posterior approach for the testing purposes calles as MAP (Maximum a Posterior approximation).
- Here we have a latency of k days while forecasting a future stock. Hence, the model becomes like- for a CHMM model λ and the values of stocks for k number of days $(O_1, O_2, ..., O_k)$ with having the values like Stocks Openprice of $(k+1)^st$ day and we are gonna find the value of closeprice of $(k+1)^st$ day which will be same as estimation of fractional change

$$\frac{Close - Open}{Open}$$

of the $(k+1)^st$ day. Then we are gonna find the MAP estimate for the Observation vector O_{k+1} on the $(k+1)^st$ day with the given values of k days.

$$O_{k+1} = \arg\max_P (O_{k+1}|O_1, O_2, \ldots, O_k, \lambda)$$

$$= \arg\max \frac{P(O_1, O_2, \ldots|\lambda)}{O_1, O_2, \ldots|\lambda)}$$

Here, as the denominator will be constant with respective of O_{k+1}, Then the MAP estimate becomes

$$O_{k+1} = \arg\max P(O_1, O_2, \ldots|\lambda)$$

- Now, we are gonna take the λ = (π,A,B) so that the probability of observations $(P(O|\lambda)$ will become local maximized to maximize the probability of observations sequences.
- We are gonna compute the probability over a discrete set of the possible values of O_{k+1}
- Now, we will use the Forward- Backward algorithms or Baum-Welch approach to compute the joint probability values of P($O_{k+1}|O_1, O_2, \ldots, O_k, \lambda$).
- Here we are gonna observe how will the sequences of π, A and B are calculated.

$$a_{ij} = \frac{\text{expected number of transitions from } S_i \text{ to } S_j}{\text{expected number of transitions from } S_i}$$

$$= \sum_{t=1}^{T-1} \xi_t(i,j) / \sum_{t=1}^{T-1} \gamma_t(i)$$

$$b_i(k) = \frac{\text{expected no of times in state } j \text{ and observing no } v_k}{\text{expected number of transitions from } S_i}$$

$$= \sum_{t=1}^{T} \gamma_t(j) / \sum_{t=1}^{T-1} \gamma_t(i) \text{ s.t. } O_t$$

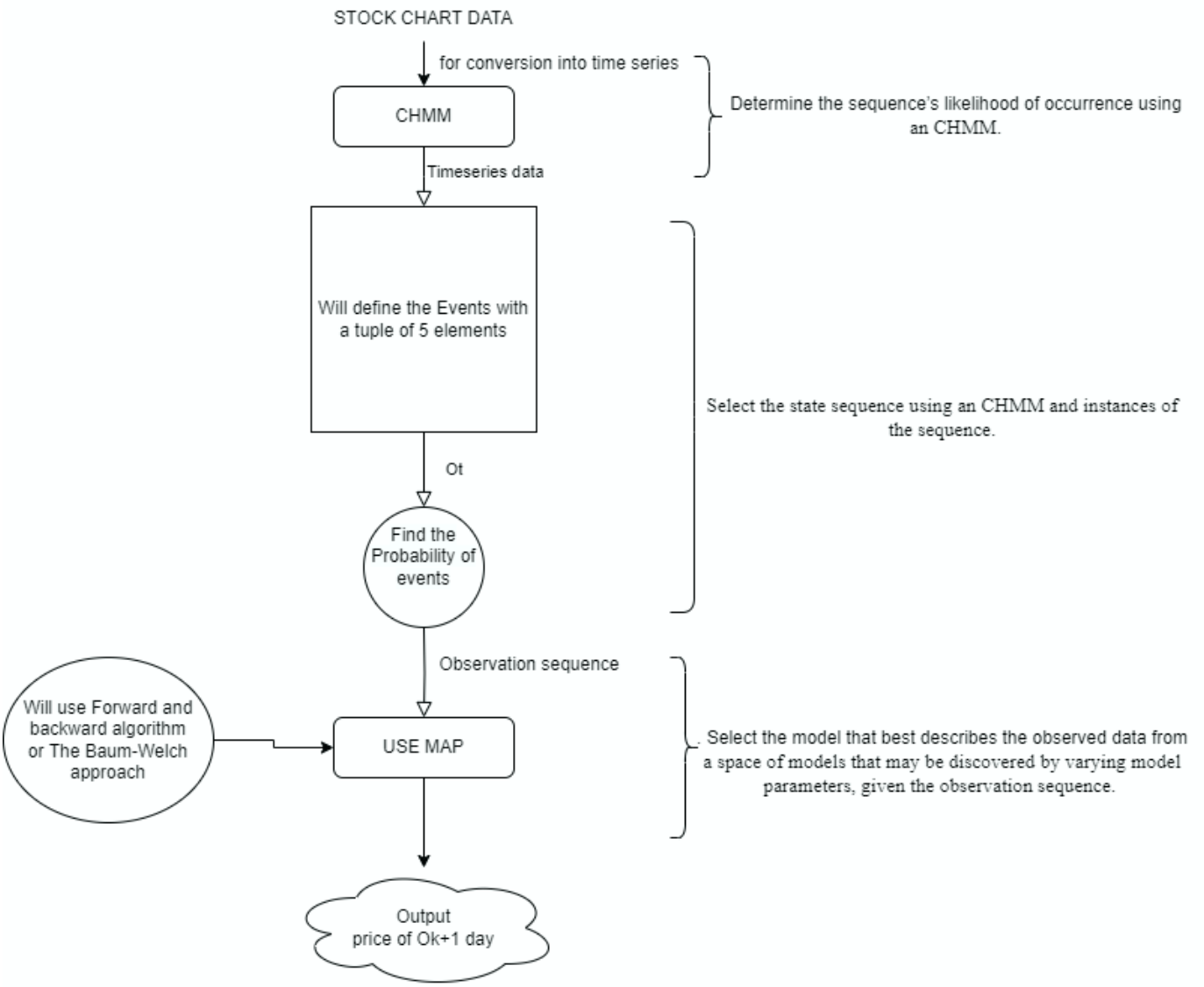

Fig. 1. FLOWCHART OF MODEL.

$$= v_k$$

- Here, $\gamma_t(i)$ is the sum of the expected number of transitions from state S_i to S_j at time tand
  π_i is the expected frequency of state S_i at a time of t=1 which is equal to $\gamma_1(i)$
- Now, after developing a CHMM , now we will have to forecast the price of the following day.
- To get this, we are going to utilizing the likelihood value from the trained CHMM and calculate the current day dataset. For example, if the price is x on a particular day, we would choose a previous instance from the past wherever the price was x or as close to x as was practical, and we would subsequently observe the behavior of that instance moving forward.
- Stated differently, we examine the stock's movement on the day when x and the price were nearly equal.
- Assuming the stock price will follow the same format as previous data on that specific value, we calculate the difference between the closing prices of that day and the next day.
- Therefore, a model to predict the stock prices have been developed.

B. THE DE BRUIJN GRAPH

In graph theory and computer science, a de Bruijn graph is a mathematical structure that is particularly useful for representing overlapping sequences and assembling sequences in bioinformatics. A multidimensional directed graph known as the De Bruijn graph with m symbols illustrates the overlaps in the symbol sequence. As a result, a sequence of symbols with length n is represented by m^n vertices in the graph.It mainly contains 4 elements nodes,edges,Eulerian path, geneome assembly.

We are gonna first employ the de Bruijn graph to the input stock chart and then use Marchov chain with probability to define the forecast stock price.

a) APPROACH :

- For instance , lets consider the following binary sequence:

0000101111010000101

- Now, lets take our n as 3 : then by definition of de bruijin's, the vertices will be : (000),(000),(000),(001),(010),(101),(011),(111),(111) ,(110),(101),(010),(100),(000)(001),(010),(101)
- For the above sequence, the de bruijin's graph will look like :

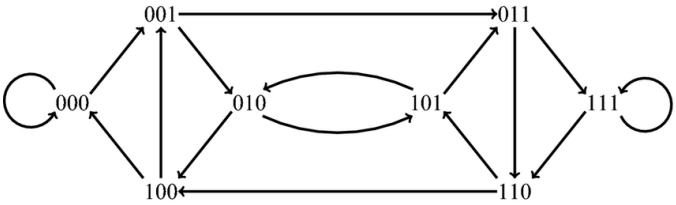

Fig. 2. De Bruijn Graph

- The edges show which piece in the sequence comes next. Based on this graph, we can infer that the series (001), (011), (111),... and so on might be the next one in the series.
- To find patterns in the data, one can use the de Bruijn graph. Once we've converted the stock chart into this directed graph and used the Markov chain to forecast market movements, we can harness the power of the de Bruijn graph.
- MARCHOV CHAIN : The Markov chain has no memory and cannot retain knowledge of previous states. It suggests that the current condition alone determines the likelihood of the subsequent change.
- We will now use the state transition matrix for the above graph

	000	001	010	011	100	101	110	111
000	0.5	0.5	0	0	0	0	0	0
001	0	0	0.5	0.5	0	0	0	0
010	0	0	0	0	0.5	0.5	0	0
011	0	0	0	0	0	0	0.6	0.4
100	0.5	0.5	0	0	0	0	0	0
101	0	0	0.4	0.6	0	0	0	0
110	0	0	0	0	0.5	0.5	0	0
111	0	0	0	0	0	0	0.5	0.5

Fig. 3. Transition Matrix

- Here these values in matrix are derived with the help of probability distribution.
- These values may vary when the length of the sequence is longer and this pattern may repeat quite often.
- Marchov Chain has a property that probability transition must have to satisfy the

$$n_i = \sum_{j \in i} n_j P_{ij}$$

- Here P_{ij} represents the state transition matrix and i, j $\in$ I (which is a State Space).
- Now for an instance, lets consider that we have the closing prices from chart for a 21 days and after that lets divide them into 3 fields : 1. High 2. Low 3. zero+, and we have these state number as 9,8,5 respectively.
- Therefore, the probability becomes,

$$P_H = \frac{9}{21}, P_L = \frac{8}{21}, P_Z = \frac{5}{21}$$

- Now,The initial vector will be ,n_0 =[0.42, 0.38, 0.23]
- Now, with the above data we are gonna calculate the transition matrix P,

$$\begin{bmatrix} 0.45 & 0.34 & 0.30 \\ 0.39 & 0.30 & 0.28 \\ 0.69 & 0 & 0.29 \end{bmatrix}$$

- Now , lets take our initial vector guess as [1 0 0]then, $n_1 = n_0.P$
  = [1 0 0] P
  n_1 = [0.45 0.34 0.30]
- Now, the marchov chain looks like :

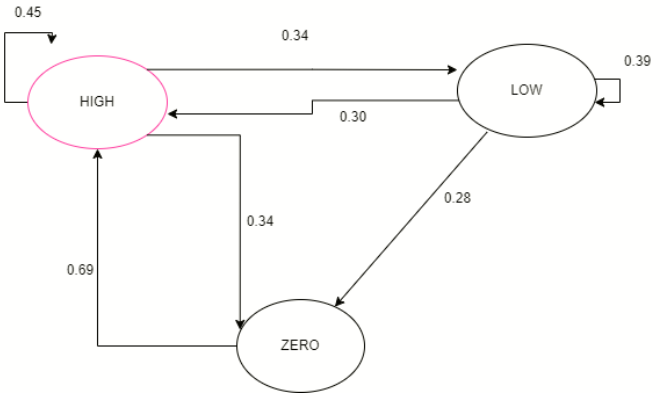

Fig. 4. MARKOV CHAIN

- It means with the given current state and the state transition distrbution , we can be able to find the next state, which implies again High.
- Therefore, we can use the above stated properties of Markov chain we can predict the stock prices for the next day and with the use of De-Bruijin Graph we can transform the given information into a graph.
- Now, for the given problem we have a array of k elements where k$\geq$2000.
- Each element in array represents the stock information of that particular day, which is basically a tuple of 5 elements : Open Price, Close price, High, Low , Volume for that day.

- A series of discrete strings and numbers serves as the foundation for the modern de Bruijn graph. But what's given to us is a list of tuples. Therefore, we must create the de Bruijn graph from this array of tuples.
- We want to know, first and foremost, what sequence of events results in a change in the stock price. In order to determine the event pattern, we compute the percentage change in the values of the five variables and "bin" them according to how much the price changed—Low (L), High (H), or Medium (M). Therefore, we have strings instead of numbers.
- Now, we will transform these events into the strings and then we can convert them into a De-Bruijin's Graph,we can assume the value of m as we want and them will have a pattern like for example, MMH to MML and the outcome which is associated with them.
- If a pattern appears that is noticeably more weighted than others using the cyclic decomposition and empirical historical data of a stock market, we can speculate that there may be some underlying determinism in the sequence that these patterns indicate.
- In order to forecast the stock price given the current state of the market, the patterns in the sequence can be further normalized and turned into a Markov chain with the use of the de Bruijn graph. The probabilities of the patterns can then be represented by the graphs' edges, which can be used to reduce the graph into an adjacency matrix, which approximates the transition matrix.
- Therefore, We have succesfully developed a model using De-Bruijn's graph and Markov chain . Marchov chains are very useful for the stock market analysis and many researchers are fond of it.

III. SIMILARITIES BETWEEN TWO STOCK CHARTS :

A. PROBLEM STATEMENT :

Given two stock charts α and β, design one or more algorithms to measure

B. SOLUTION APPROACHES :

The algorithm known as Dynamic Time Warping (DTW) can be used to match two time series that are subject to temporal shifts. Time series data analysis is one of its common uses. In many scientific domains, data that can only be analyzed after accounting for time represent a significant problem. It is employed in the processing of financial, meteorological, and ECG data.(please refer figure[5])

a) APPROACH :

- Firstly, we have two stock charts α and β.
- Now, we will convert the stock data from both charts α and β into time series by using Inverse Wavelet Transform which will change the frequency series into times series or with the method that has stated above using CHMM.
- Then now,we will have 2 time series lets represent them as T_α and T_β
- Now, we will pass the time series into DTW algorithm which will be used to measure the similarity between the two time series
- We will have to pass a constrain in form of a distance d which is the distance to the diagonal matrix. We will use Sakow-Chiba Constrain for this.
- For example, if we use d = 10 it means that the pattern of T_α can be connected with the same pattern of T_β which has occured upto 10 hours earlier or later.
- Now , we will use the cross-correlation to measure the similarity.
- It is calculated using this formulae:

$$r(T_\alpha, T_\beta) = \frac{(T_\alpha - \bar{T_\alpha}).(T_\beta - \bar{T_\beta})}{\|(T_\alpha - \bar{T_\alpha}\|_2 \|(T_\beta - \bar{T_\beta}\|_2}$$

- We will get the value of r which is the measure of similarity between the two stock charts α and β.

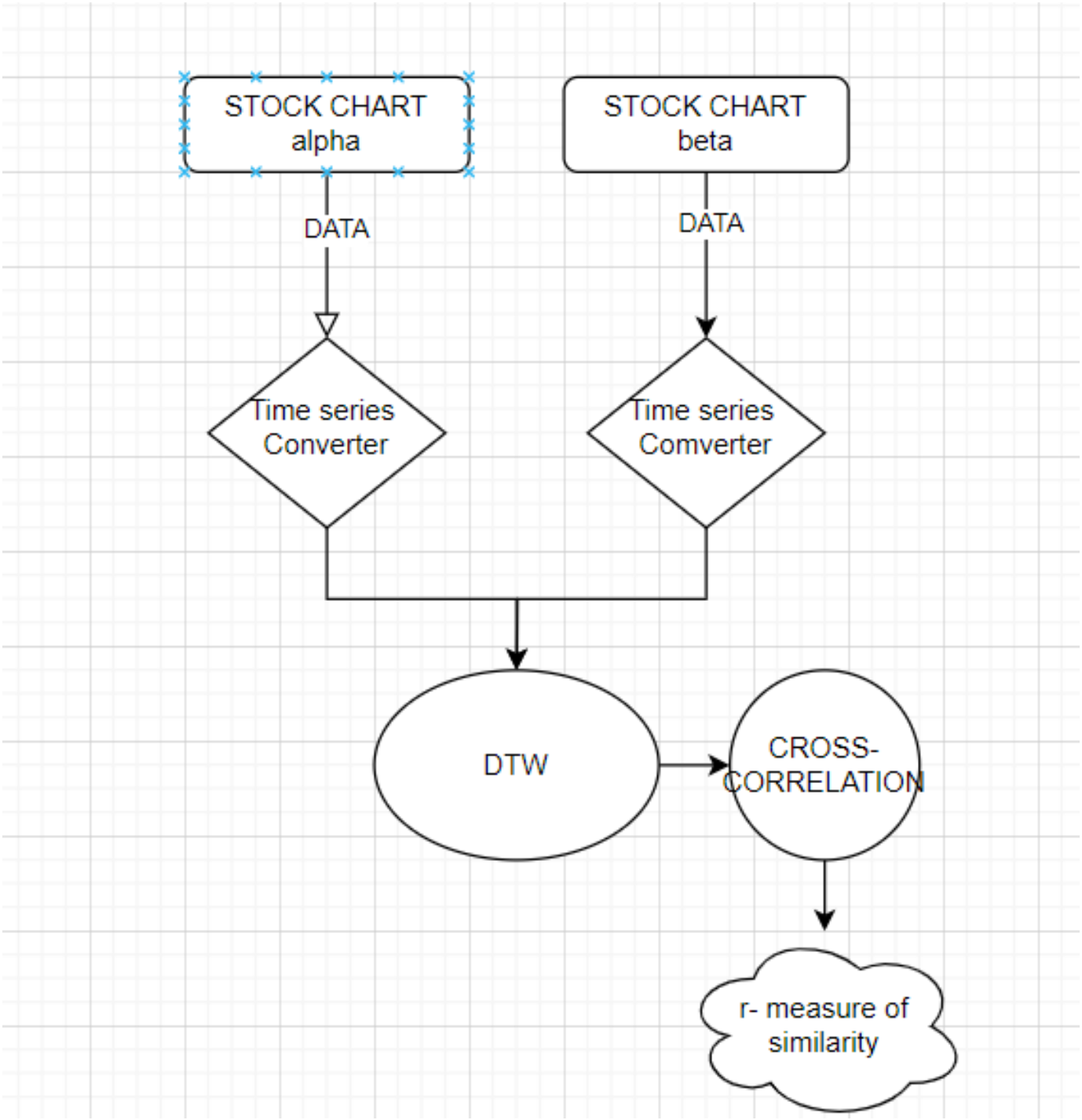

Fig. 5. Dynamic Time Wraping

1) USING KRUSKAL's ALGORITHM or PRIM's ALGORITHM :

* Now we are gonna use a MST technique to develop a stock network in which we will have nodes as the stock prices for the given stock charts, α and β
* Here, the edges will act as separators between the two stocks.
* Then we are gonna employ Krushkal algorithm or Prim's Algorithm to it.

a) KRUSKAL's ALGORITHM :

* Sort Edges: Put each edge in the weights' non-descending order.
* Initializing MST : Set up an empty graph to symbolize the least spaning tree in order to initialize MST.
* Repeat Along Edges: Add edges to the MST one at a time, avoiding cycles, starting with the smallest edge. You should omit an edge if adding it starts a cycle.
* Proceed with Iteration: Until the MST has (n-1) edges, repeat step 3 until n, the number of vertices in the original graph, is reached.
* Output MST: The minimal spanning tree is the graph that is produced.

b) PRIM's ALGORITHM :

* Start with an arbitrary vertex: Select a random beginning vertex.
* Set up MST: To represent the lowest spanning tree, make an empty graph.
* Choose the Edge That Is the Smallest: To connect a vertex inside the MST to a vertex outside the MST, choose the smallest edge possible.
* Include Vertex in MST: To the MST, add the chosen edge and its matching vertex.
* Repeat Cycle : Until every vertex is included in the MST, repeat steps 3–4 once more.

RESULT: Here we are gonna observe MST cycles, if there is any resembelence and correlation between them they will run the cycles. If they are not same then the cycles does not runs. Therefore, if cycles runs our α and β are identical otherwise, they are not identical.

2) USING RECTANGULAR GRANULATION INFORMATION :

* Firstly, we are gonna convert the frequency series into data series then the methodology follows for the estimation of similarity between stock charts α and β.
* Rectangular information granulation of time series aims to produce specific and justified rectangular information granules in the 2-D plane consisting of time series data and its first order difference time series. For it to satisfy the conditions, two things must happen.
* Justifiability: Inside the parameters of the generated information granule, as much original data as is practical must be gathered. It increases the value of the information. Specificity: Since each generated information granule should have a clear meaning, the smaller the information granule, the better.
* We will now build a two-dimensional plane by Y_A and Y_B, where the time series is represented as a set of points (yi, yi).Building a rectangular information granuale on a plane that satisfies the previously stated justifiability is the primary goal. In this case, the restrictions over Y_A and Y_B are given the median, min, and max. Granulating regular information is an optimization problem. Right now its four decision factors are used to create information granules using the particle swarm optimization algorithm.However, Gravitational Search Algorithm(GSA) produces more accurate information granules than the swarm optimization method since it has a faster convergence time and better global search capabilities.For both lower and upper boundaries, the GSA is employed.Here, several information granules are counted to determine changes in the time series. Next, granule structure is determined using FCM, and the granules are formed into different C clusters to produce a partition matrix.
* Calculating Similarity : In contrast to the traditonal method, which involves calculating the distance between two time series and then calculating the similarity between them, this method measures the similarity between time series within the context of rectangular information granulation.(please refer the Figure[6].

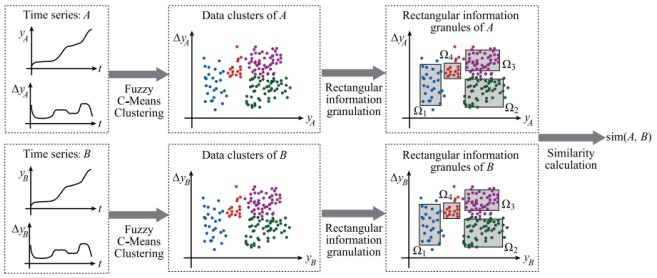

Fig. 6. RECTANGULAR INFORMATION GRANULATION

IV. CONCLUSION

We have employed various models that can be helpful for us to predict the forecast of stock prices and similarity between two stock charts. For every

model, it is essential to firstly convert them into time series from the frequency series given, then we can use many various methods to determine the given queries. We have used the power of Markov Chains for the transition of the data into time series. Then we have employed CHMM's, DTW, De-bruijn graph etc to get the desired outcome.In the second question , I rephrased the query to similarity between two times series as we have already mentioned ways for the conversion into the time series. Therefore, we have successfully achieved the desired models that can be used to solve the given queries with input being only the stock charts.

V. FUTURE WORK :

We can solve this by using many other methods like Fast Fourier Transform, Genome Sequence Analysis,..etc. There are many other neural network techniques and machine learning techniques that are used now a days for the stock market analysis but with the given input of length which is $k \geq 2000$ it hard to equip machine learning techniques into our required model as most of them uses LSTM and Short term analysis for the processing. We need even more data to achieve them. So, we have only used the conventional mathematics and computer science technologies for the designing of our models.

VI. ACKNOWLEDGEMENT

I hereby, respectfully thank Professor Dr.Zhe Dhang for his contribution to this course and giving us a strong foundation on the advannced algorithms and making us to think out of the box. I am dedicating this work to him.

REFERENCES

[1]Markov Mode for Stock Trading by Nguyet Nguyen https://www.mdpi.com/2227-7072/6/2/36

[2]Research on Marketing Prediction Model Based on Markov Prediction by Haiying Chen,1Haiyan Chen,2Wei Zhang,1Chaodan Yang https://www.hindawi.com/journals/wcmc/2021/4535181/

[3]Baum-Welch algorithm for training a Hidden Markov Model — Part 2 of the HMM series https://medium.com/analytics-vidhya/baum-welch-algorithm-for-training-a-hidden-markov-model-part-2-of-the-hmm-series-d0e393b4fb86

[4]Information Granulation With Rectangular Information Granules and Its Application in Time-Series Similarity Measurement Sheng Du , Graduate Student Member, IEEE, Min Wu , Fellow, IEEE, Luefeng Chen , Member, IEEE, Weihua Cao , Member, IEEE, and Witold Pedrycz , Life Fellow, IEEE

[5]Dynamic Time Warping in Financial Data – Modification of Algorithm in Context of Stock Market Similarity AnalysisTomasz Grzejszczak1 , Eryka Probierz1,2, Adam Gałuszka1 , Krzysztof Simek1 , Karol Jedrasiak3 , Tomasz Wiśniewski4

[6]Offline and Online Identification of Hidden Semi-Markov ModelsMehran Azimi, Panos Nasiopoulos, and Rabab Kreidieh Ward, Fellow, IEEE

[7]"Hidden markov models and the baum-welch algorithm". IEEE Information theory society newsletter, Dec 2003.

[8]B. Nobakht, C.E.J. Dippel, and B. Loni. "Stock market analysis and prediction using hidden markov models", unpublished.

[9]M.R. Hassan. "A combination of hidden markov model and fuzzy model for stock market forecasting". Journal of Neurocomputing, pages 3439– 3446, 2009.